



Why Laptops and Other Edge Devices are Being AI-Enabled

In this handbook:

Why Laptops and Other
Edge Devices are Being
AI-Enabled

Why Laptops and Other Edge Devices are Being AI-Enabled

BRYAN BETTS, PRINCIPAL ANALYST

Why do we need dedicated AI engines in laptops, and especially in the ultrathin laptops that AMD, Qualcomm and Intel are all targeting with their latest processors? What's the motivation behind Intel's AI Everywhere and AI-PC programmes, and where is the desktop PC in all this?

First, it's important to remember that "AI" is much, much more than just chatbots, co-pilots and other intelligent assistants. It is a catch-all that covers a very broad spread of technologies, some of which have been in widespread use for many years.

AI TRAINING AND INFERENCING HAVE DIFFERENT WORKLOAD PROFILES

It's also essential to recognise that the vast majority of AI "workloads" involve inferencing, which is the process of applying a trained model to a particular situation. And as more and more personal and office applications incorporate AI techniques, this is where AI-enabled laptops come in, along with the AI-enabled smartphones and other mobile devices that are also coming down the line.

In this handbook:

Why Laptops and Other
Edge Devices are Being
AI-Enabled

Inferencing is a complex task, but nowhere near as heavyweight as the job of training that model in the first place. Training is still the role of the hundreds and thousands of high-end GPUs from Nvidia and others that fill servers world-wide, and of course all the other specialist AI chips that currently sit inside AWS, Azure and Google data centres.

Personal productivity examples of using AI inferencing – or rather, its ML subset – include facial recognition for FaceID, and improving the real-time background blur and eye-tracking within videoconferencing applications. There's many other areas where inferencing can greatly improve performance, though, including graphic design, photo and video processing, 3D modelling, AR/VR, data science and much more.

WHERE TO RUN YOUR AI INFERENCING

Of course, there's several ways to run these workloads. You can run them in the cloud, but as well as the inevitable latency this involves, it's also increasingly costly both in terms of network bandwidth and cloud compute costs. There's also the governance issue of sending all that potentially-sensitive and bulky data to and fro.

So at the very least, doing a first-cut and filter to reduce and/or sanitise the transmitted data volume, is valuable in all sorts of ways. You could use the GPU or

In this handbook:

Why Laptops and Other
Edge Devices are Being
AI-Enabled

even the CPU to do this filtering, and indeed that's what some edge devices will be doing today.

Alternatively you could simply run the inferencing work on the local CPU or GPU in your laptop or desktop. That works, but it's slower. Not only can dedicated AI hardware such as an NPU do the job much faster, it will also be much more power-efficient. GPUs and CPUs doing this sort of work tend to run very hot, as evidenced by the big heatsinks and fans on high-end GPUs.

AI ON THE DESKTOP TOO?

That power-efficiency is useful in a desktop machine, but is much more valuable when you're running an ultraportable on battery, yet you still want AI-enhanced videoconferencing, speedy photo editing, or smoother gaming and AR.

That's a big part of why all these companies, and Apple too, target ultraportables first for their newest AI-capable processors. Indeed, on the Windows side of things, it's pretty much all laptops (and servers, via Intel's AI-enabled 5th gen Xeon processors) for now, with desktop NPU-equipped chips promised for later.

In this handbook:

Why Laptops and Other Edge Devices are Being AI-Enabled

Apple has had AI-enabled Mac Mini and Mac Studio desktops for quite a while now, though, and users we've spoken with say the application performance boost over non-AI processors can be massive, with multi-tasking also significantly improved.

THE BUYING DECISION: WHAT AND WHEN TO CHOOSE

One caveat to draw from all this is that there's no obvious "killer app" for on-board AI. However, it enables all sorts of new capabilities and power boosts that are individually relatively small, but which when added together can massively enhance productivity and user experience.

Much will also depend on software support, of course, which with these latest processors and systems that means Microsoft Windows and its application ecosystem, but the lessons are already there in the Mac world. All in all, they say that a processor with a dedicated AI engine or core is going to become essential as more and more software takes advantage of AI capabilities and technologies.

So if you're involved in specifying desktops and laptops for your organisation, start by identifying the groups of users most likely to benefit from these new AI-enabled systems. Many will be power users currently equipped with workstations or high-end laptops.

In this handbook:

Why Laptops and Other Edge Devices are Being AI-Enabled

For example, this might include groups such as engineers and scientists using 3D modelling and visualisation software, financial and business analysts, data scientists, and creatives working with sound, images and video – though of course the latter may already be AI-enabled from having taken the Apple route. All these people are likely to be both well-paid and able to get more done with a faster machine, so there's a strong business case here.

Read more AI & ML content from Freeform Dynamics [here](#).